General comments:

Upon revision of the manuscript, we noticed a small number of duplicates in the experimental data sets that we have crawled from pubchem (the “broad” and “confined” data sets), some of which had wrong vapor pressures. This was a result of an error upon merging data after crawling. We removed all duplicates, assembled new data sets and repeated all calculations. Accordingly, we have updated the majority of figures, as well as various numbers and information across the manuscript. In summary, due to the small number of duplicates, no major changes occurred and the general observations and conclusions remain unchanged. The published data sets and models have been updated accordingly.

In addition, we have revised the learning curve experiment, which we had only conducted for fdGC^{2}NN models in the previously submitted version. We have now generated learning curves for the adGC^{2}NN models as well. When obtaining this learning curve, we did not conduct any data set size-specific hyper-parameter tuning for the fdGC^{2}NN and adGC^{2}NN models. This way, the learning curves are representative for the model architectures and hyper-parameters that we provide in the manuscript.

Reviewer 1:

General comments:

This manuscript presents a compelling machine learning framework—group contribution-assisted graph convolutional neural networks (GC2NN)—for improving vapor pressure predictions of organic and inorganic molecules. The study is comprehensive, technically sound, and well-articulated. It provides a rigorous benchmark against established methods and convincingly demonstrates the advantages of adaptive-depth GC2NN models, especially in handling compounds with limited experimental data. The paper is suitable for publication pending minor revisions to improve clarity, reproducibility, and contextualization of the results.

We would like to thank reviewer 1 for taking the time to review our manuscript and highly appreciate the positive and constructive feedback. The questions, thoughts and ideas are very helpful and we implemented the suggested changes where possible.

We will address the reviewer's (R1) comments (C1-C10) in point-by-point answers below. Suggested changes to the manuscript are intended and given in *italic font*.

R1C1. Abstract: Briefly explain what "group contribution-assisted" means in lay terms.

We would like to thank reviewer 1 for this proposition, and agree, that readers of the manuscript would benefit from an easy-to-understand introductory sentence to this concept. We have therefore added a sentence to the abstract.

> *l. 8 – "The model's group contribution component is a shallow fully-connected neural network which processes numerical molecular descriptors and complements the model's graph component."*

R1C2. Model Hyperparameters: Include a summary of final selected hyperparameters in the main text (not only supplementary).

We agree that a brief summary of the final hyperparameters of the adGC^{2}NN models would be relevant for readers of the manuscript to be summarized in the main text, and have added this summary to the model architecture and training description in the results section:

> l. 197 – *"All adGC^{2}NN models possess two hidden layers with each 32 nodes in the group contribution component and a single merging layer with eight nodes for each graph convolution layer. The graph component of the adGC^{2}NN models is comprised of a total of five layers with 32, 16, 64, 16 and 32 nodes, using 'LeakyReLU', 'LeakyReLU', 'ReLU', 'ReLU' and 'LeakyReLU' activation functions, respectively. Among these, the second and fifth layers are graph attention layers with six attention heads each, processing additional edge information. Training is conducted with a learning rate of 6.25x10^{-4}, a learning rate decay of 0.985 per training epoch, no weight decay and a batch size of four (Fig. 2, Tab. S6)."*

R1C3. Loss Function Choice: Explain why MAE was chosen over RMSE or other metrics, especially given outlier sensitivity.

We are grateful for this suggestion. We have added the information to the model architecture and training description in the methods section:

> l. 163 – *"We select mean absolute error (MAE) as loss function for model training, as well as model evaluation and comparison with established methods, due to its robustness over methods that give more weight to outliers such as root mean squared error (RMSE). This is particularly important given that the training data consist of experimental measurements that may possess high uncertainty and could be subject to systematic biases originating from different experimental setups. Measurements in the ELVOC range are particularly susceptible to higher experimental uncertainties, which would receive disproportionate weighting under RMSE-based training and consequently degrade model performance on other ranges. MAE allows for a reliable and interpretable evaluation of model accuracy without being overly influenced by extreme values."*

R1C4. Model Training Time: Mention hardware specifications (already noted) in the main text to aid reproducibility.

As the reviewer indicated, we have already noted the GPU hardware specifications for the model training. To add clarity for readers, we have added more hardware as well as software specifications to the model architecture and training description in the methods section:

> l. 177 – *"Hyperparameter optimization and model training are conducted on the Raven high-performance computing (HPC) system of Max Planck Computing and Data Facility (MPCDF), which provides GPU-accelerated compute nodes, each with four Nvidia A100-SXM4 GPUs and 160 GB HBM2. Each model is trained on a single Nvidia A100-SXM4 GPU using up to 24 GB of memory and PyTorch version 2.4.0 with CUDA version 12.1 support."*

5. **模型可解**释性：该模型能否提供对哪些官能团影响最大的 pvap **的**见解？欢迎任何特征归因分析。

6. **不确定性分析**：**考**虑量化 MAE **之外的不确定性**（**例如**，**集成**预测或贝叶斯 GCNN）**以解决**实验变异性。

7. ELVOC **挑**战：扩展潜在策略，以提高 ELVOC **模型的准确性**，**而不**仅仅是数据集扩展。

8. **大气建模相关性**：**更明确地强**调您的结果如何改进 SOA **模型或区域气候模**拟的参数化。

9. **延伸**电位：这种方法是否可以适用于其他物理化学性质（例如，亨利定律常数、反应性）？对此添加一句话。

10. 结论可以简要提及集成物理信息神经网络或混合 QM/ML **模型的潜在未来**发展。

We will respond to the following English translations of the reviewer comments:

R1C5. Model interpretability: Can the model provide insights into which functional groups have the greatest impact on pvap? Any feature attribution analysis is welcome.

We thank reviewer 1 for this comment that prompted us to look further into the topic of feature attribution. The utilization of graph attention layers in our model provides a straightforward way for feature attribution analysis and we now calculated attention scores based on the test sets. Accordingly, we have added several sections across the manuscript to better explain the GAT layers and how they can be used for feature attribution analysis.

> l. 150 – *"Graph attention layers utilize attention mechanisms, enabling them to weigh convoluted nodes and features by their importance (Velickovic et al., 2017; Withnall et al., 2020; Tang et al., 2020). This capability allows the assessment of feature importances by evaluating attention weights (Sanchez-Lengeling et al., 2020). Furthermore, graph attention layers enable the model to also derive information from edge attributes (Battaglia et al., 2018)."*
>
> l. 221 – *"For a general feature attribution analysis, we investigate attention scores of the second layer (graph attention) of the trained model's graph component. The attention weights, which are trained parameters of the model, are applied to each chemical compound and graph node (i.e., atom) to compute attention scores. They represent the calculated importances, quantifying the contribution of each node to the p_{vap} prediction relative to its neighboring nodes for a specific compound. For functional groups, importances of all associated atoms are averaged. With regards to single atoms, oxygen (0.36) scores a slightly larger attention score than carbon (0.32) in the confined test set (Fig. S5). Among functional groups, hydroxyl groups achieve the highest score."*
>
> l. 270 – *"Calculated attention scores for single atoms and functional groups are summarized in Fig. S11. Notably, we observe a good agreement between the attention score orders of functional groups between adGC^{2}NN-confined and adGC^{2}NN-broad, with hydroxyl groups having the highest scores, followed by carbonyl groups, ester groups and finally non-aromatic C=C double bonds. The importance of hydroxyl groups may be attributed to their ability to form hydrogen bonds that reduce the compound's vapor pressure. Note that feature importances assigned by trained models are not exclusively governed by chemical principles, but also the prevalence and distribution of substructures in the training data. Rarity and commonness of certain substructures may both decrease associated feature importances, as high importances are attributed to relevant features that enable the model to distinguish compounds of the training population. To differentiate between chemistry-governed and prevalence-governed importances, feature attribution analyses could be supported by generative sensitivity studies, where the effect of substructures on p_{vap} predictions is statistically tested through systematic substitution of substructures in template compounds."*

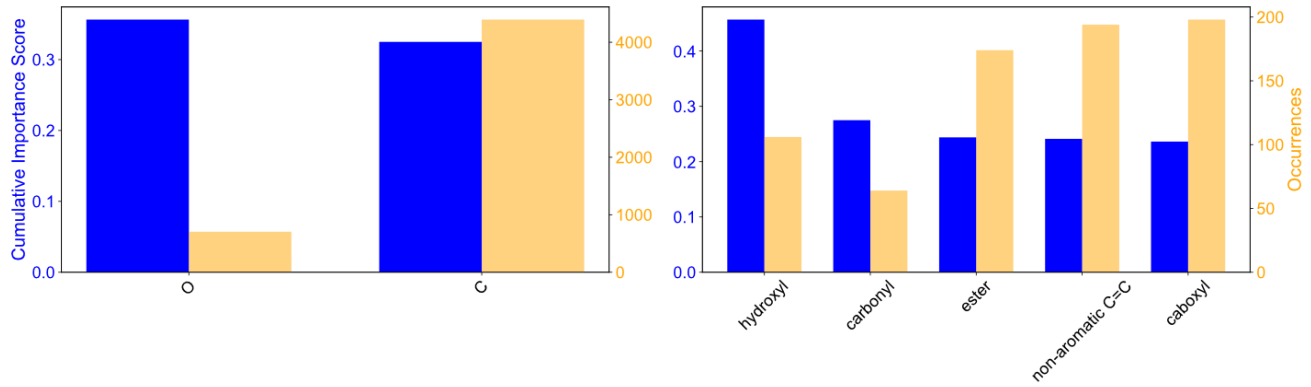

Figure S5. *Cumulative importance scores and occurrences of elements and functional groups in the confined test set (organic compounds with a limited set of functional groups), calculated in the second layer (graph attention layer) in the graph component of the trained T+V adGC²NN-confined. Specifically, self-loop importances of the nodes attributed to various elements or functional groups are averaged to determine their relative importance among all neighboring nodes they are convoluted with.*

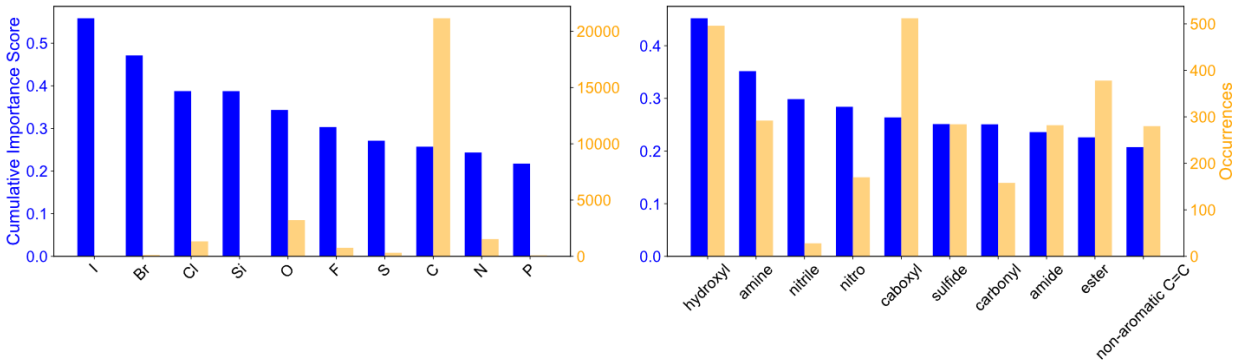

Figure S11. *Cumulative importance scores and occurrences of elements and functional groups in the broad test set (including inorganic compounds), calculated in the second layer (graph attention layer) in the graph component of the trained T+V adGC²NN-broad. Specifically, self-loop importances of the nodes attributed to various elements or functional groups are averaged to determine their relative importance among all neighboring nodes they are convoluted with.*

R1C6. Uncertainty analysis: Consider quantifying uncertainty beyond MAE (e.g., ensemble prediction or Bayesian GCNN) to account for experimental variability.

We appreciate this suggestion and have analyzed ensemble predictions of our five adGC²NN cross-validation models for the confined and broad data set with regards to ensemble mean error and ensemble standard deviation serving as an indicator of model uncertainty. The plots added to the SI (Figs. S6 and S10) give an overview of molecular structures in the test data where model error or model uncertainty are particularly high. We have also added the following statements to the main text of the manuscript:

l. 232 – *"To assess model uncertainty, we analyze ensemble predictions from the 5-fold cross-validation models on both confined and broad test data sets with regards to their prediction errors and standard deviations (Fig. S6). While the ensemble mean error represents model bias, the ensemble standard deviation can serve as an indicator for overall model uncertainty."*

l. 268 – *"An uncertainty analysis based on ensemble predictions for the broad test data of the 5-fold cross validation adGC^{2}NN-broad models is shown in Fig. S10."*

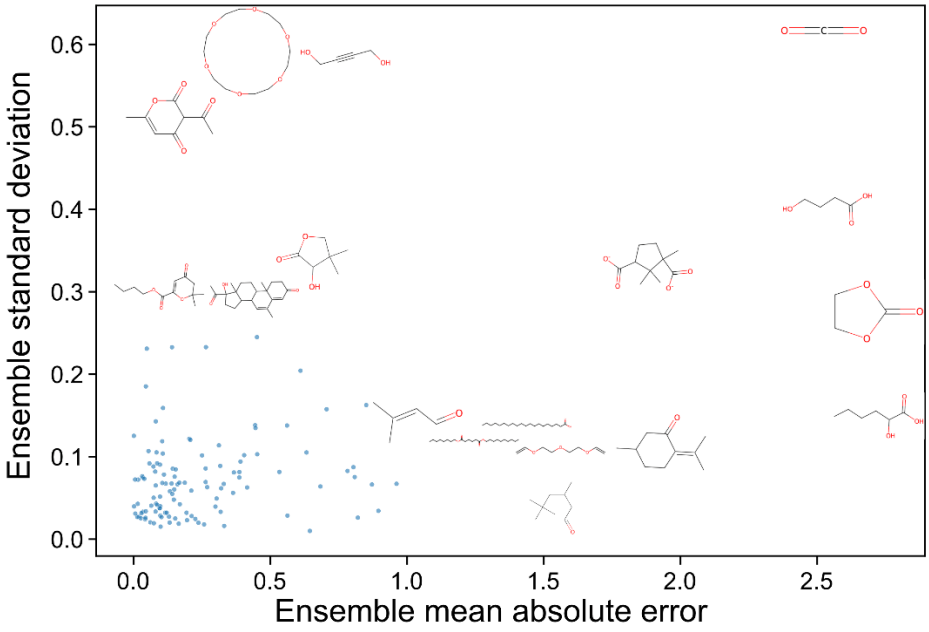

Figure S6. *Ensemble standard deviation as a function of ensemble mean absolute error for the confined test set. Ensemble predictions originate from the confined adGC^{2}NN 5-fold cross validation models which are trained on different subsets of the training data. All compounds with an ensemble standard deviation larger than 0.3 or an ensemble mean absolute error larger than 1.0 are plotted as molecular structures. The compounds on the top of the figure are associated with large model uncertainty, while compounds in the bottom right have large errors despite small model uncertainty, a potential indicator for experimental uncertainty.*

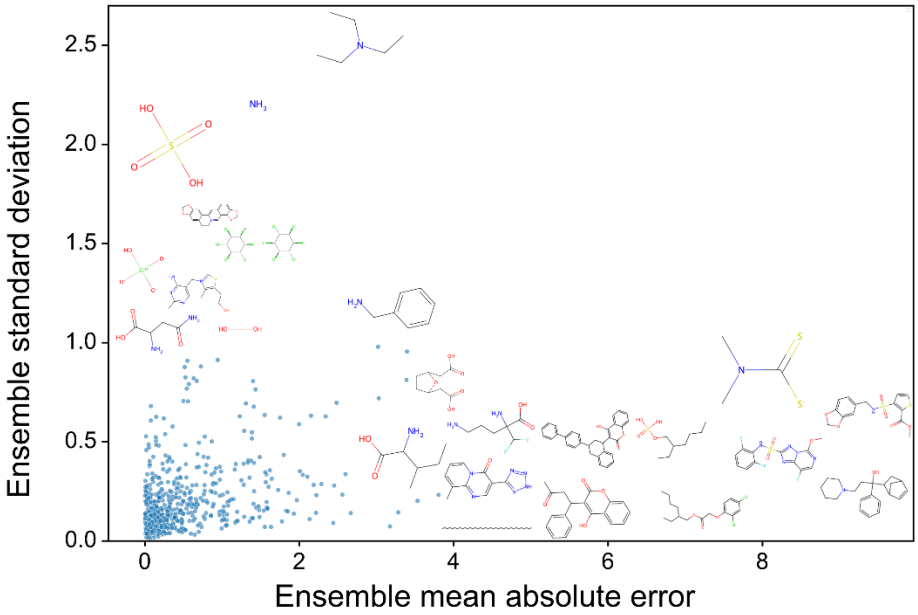

Figure S10. *Ensemble standard deviation as a function of ensemble mean absolute error for the broad test set. Ensemble predictions originate from the broad adGC2NN 5-fold cross validation models which are trained on different subsets of the training data. All compounds with an ensemble standard deviation larger than 1.0 or an ensemble mean absolute error larger than 4.0 are plotted as molecular structures. The compounds on the top of the figure are associated with large model uncertainty, while compounds in the bottom right have large errors despite small model uncertainty, a potential indicator for experimental uncertainty.*

R1C7. ELVOC Challenge: Scaling potential strategies to improve the accuracy of ELVOC models beyond just dataset expansion.

If we understand this comment correctly, reviewer 1 suggested to discuss potential strategies for dealing with data scarcity in the extremely low volatile organic compound (ELVOC) range. We have added a few ideas to the conclusion section:

> l. 351 – *"The problem of data scarcity is very evident for compounds in the ELVOC range, which are comparably rare and underrepresented in our data set. This may be due to greater difficulties in the experimental determination of saturation vapor pressures of ELVOCs. To accurately extend QSAR models to the ELVOC range, possible strategies may include the utilization of quantum mechanical-derived data instead of experimental data, or potentially the application of more advanced machine learning models that include heuristic rules or physics-informed modules (Bilde et al., 2015), transfer learning to enable extrapolation outside of the training domain (Lansford et al., 2023) or pre-trained models that can be fine-tuned using small data sets (Burns et al. 2025)."*

R1C8. Atmospheric modeling relevance: Emphasize more explicitly how your results improve the parameterization of SOA models or regional climate simulations.

We have added a small paragraph clearly summarizing our contribution with regards to the data sets as well as trained models that are published along with our manuscript:

> l. 378 – *"The data sets (broad and confined) as well as the associated trained models are published along with this study. The compiled experimental vapor pressure data can be used for future benchmarking or training of vapor pressure estimation methods. Furthermore, our trained adGC^{2}NN models can be downloaded as easy-to-use executables, enabling researchers in various fields to obtain accurate vapor pressure predictions for their research, e.g., in the fields of SOA modeling or climate simulations. To run the models, no knowledge on machine learning or programming is required."*

R1C9. Extension potential: Can this method be applied to other physicochemical properties (e.g., Henry's law constant, reactivity)? Add a sentence to this.

Yes, this is the case. Although we have discussed the potential of models which predict multiple properties at a time, we so far did not mention that the model could in principle be applied to other properties as well. We have added a sentence on this, as suggested:

> l. 357 – *"Our adaptive-depth model, however, achieved overall good results given relatively few training data, making the architecture a promising candidate for QSAR models addressing other molecular properties with relevance for atmospheric chemistry and physics, such as Henry's law solubility coefficients or reaction rate coefficients."*

R1C10. The conclusion could briefly mention potential future developments of integrated physics-informed neural networks or hybrid QM/ML models.

We agree that this concept could and should be mentioned and have added the following sentence to the conclusions:

> l. 364 – *"In a similar fashion, QSAR models can be improved through integration of physics-informed models or hybrid quantum-mechanical/machine learning models (Zhang et al. 2018)."*

Reviewer 2:

General comments:

The manuscript by Krüger et al introduces group contribution-assisted graph convolutional neural networks (GC2NN) as a novel machine learning approach to predict the vapor pressures of organic molecules, which are critical for understanding secondary organic aerosol (SOA) formation. By combining molecular descriptors with graph-based representations of molecular structure, the adaptive-depth GC2NN models significantly outperform traditional methods, especially in data scarce regimes. Two models were developed: a general-purpose model with a mean absolute error (MAE) of 0.67 log-units and a specialized model for SOA-relevant compounds achieving a much lower MAE of 0.36 log-units, demonstrating high predictive accuracy.

The manuscript is clearly written, the approach is sound and the findings are well described. Graph convolutional neural networks are relatively new in molecular atmospheric science and the finding will be of interest to the readers of GMD. I recommend the manuscript for publication in GMD provided my minor comments below will be addressed.

We would like to thank reviewer 2 for taking the time to review our manuscript and highly appreciate the positive and constructive feedback. The questions, thoughts and ideas are very helpful and we implemented the suggested changes where possible.

We will address the reviewer's (R2) comments (black, C1-C9) in point-by-point answers (blue) below. Suggested changes to the manuscript are intended and given in *italic font*.

R2C1. "We assembled a data set of SMILES representations of 6128 compounds with experimental saturation vapor pressure measurements... " When I read this sentence, I was wondering, if the dataset would be made available. Later, in the data availability statement, it is clarified that the data can be found alongside the code. It might be helpful to already allude at this point in the manuscript to the data availability.

The data set will be made publicly available. We agree that this should be alluded to more clearly in the main text of the manuscript. We have added the following statement at the end of the paragraph where experimental vapor pressure data are presented:

l. 110 – "Both data sets are available for download, as specified in the data availability statement."

R2C2. Figure 1. It is interesting, that the vapour pressure distribution of the confined dataset (shown in panel d) is skewed towards higher vapour pressures. Could this be explained and could this explanation be added to the manuscript?

Yes, the difference in vapor pressure distribution can – and should - be easily explained through a difference in molecule size. We have added the following statement:

l. 108 – "This skew towards higher vapor pressures in the confined data can be attributed to smaller molecules that contain fewer heavy atoms, as indicated by its lower average molecular mass of 154.8 g mol^{-1}, compared to 205.8 g mol^{-1} in the broad data set."

R2C3. GC2: What exactly is the input to the graph convolutions? Figure two is scant on details in this regard.

We appreciate the comment and agree that additional information on the inputs and functionality of graph convolutional neural networks would be helpful for readers, especially those unfamiliar with machine learning methods. We have added a short description of the inputs of each layer and how they are modified throughout the GC^{2}NN.

l. 133 – "Graph convolution layers receive the adjacency matrix indicating which nodes (i.e., atoms) are connected, as well as the node feature matrix as inputs, graph attention layers additionally receive edge features. While the adjacency matrix remains unmodified to allow deduction of the connectivity for the following layers, each graph layer alters the feature matrix or matrices by aggregating features from neighboring nodes or edges, using the adjacency matrix to guide the aggregation."

R2C4. I am trying to understand the added value of the group contribution component. The functional groups are already part of the graph. So in principle, one would assume that nothing new is added by supplying them separately as input. Could we think of this extra channel as some kind of weighting that increases the importance of the functional groups in the input representation? One would think that the neural network would adjust such a weighting internally already during the training, just from graph input, if it deems functional group features to be particularly important. But maybe suppling the group contribution separately enforces a higher importance from the start.

We had already elaborated on the difficulties of pure graph convolutional neural networks in the discussion section. For a better understanding, we have refined and extended this paragraph which now reads as follows:

l. 305 – "In addition to the adGC^{2}NN and fdGC^{2}NN models, we tested graph-only models without the additional input layer to obtain holistic molecular information (group-contribution component). These pure GCNN models are associated with significantly larger errors for nearly all data sets and sizes (Fig. S13), despite data set size-specific hyper-parameter tuning. This can be attributed to graph convolutions which, in principle, are merely a succession of local operations on subgraphs. In other words, a pure graph convolutional neural network performs local operations on the input graph that are independent and unaware of operations and interpretations that occur in distant areas of the molecular graph. Deducting and learning holistic molecular information only from local convolutions on the graph structure is difficult, especially for the larger molecules. As each additional convolution layer increases the distance allowed for two nodes (and hence atoms) to influence each other, setting the number of graph convolution layers to the largest distance between two nodes in the data set would enable the model to derive information from each molecule as a whole. However, intramolecular interactions are usually not long ranged. Furthermore, this is detrimental for most model training because it would result in very deep neural networks which would likely over-fit on most data sets. Therefore, since the graph neural

network training might not effectively capture whole-molecule properties, the lack of information on general molecular properties, like molar mass, inhibits the graph-only models to generalize between molecules of different size. We observe that the addition of molar mass as an input is crucial for the performance of GC^{2}NN, while additional descriptors like element and functional group counts lead to further, but minor improvements."

R2C5. I would appreciate more details on the GC2 architecture (since these are also not supplied in the SI). How exactly is the merging done in the merging layers?

We would like to refer the reviewer to the response to comment R2C3, where we have attempted to clarify the functionality and intermediate data structures of the graph component of the GC^{2}NN. Furthermore, we have expanded the explanation of the merging in the model architecture section:

Fig. 2 caption – "After passing all graph layers applicable to a compound, the convoluted and flattened node and edge feature matrices are concatenated with the processed data from the group contribution component. Fully-connected merging layers process these vectors and map them to the single-node output layer, the p_{vap} prediction."

Finally, as response to comment R1C2 we have added the selected hyper parameters for the adGC^{2}NN models to the main text, providing further information regarding the model architecture:

l. 197 – "All adGC^{2}NN models possess two hidden layers with each 32 nodes in the group contribution component and a single merging layer with eight nodes for each graph convolution layer. The graph component of the adGC^{2}NN models is comprised of a total of five layers with 32, 16, 64, 16 and 32 nodes, using 'LeakyReLU', 'LeakyReLU', 'ReLU', 'ReLU' and 'LeakyReLU' activation functions, respectively. Among these, the second and fifth layers are graph attention layers with six attention heads each, processing additional edge information. Training is conducted with a learning rate of 6.25x10^{-4}, a learning rate decay of 0.985 per training epoch, no weight decay and a batch size of four (Fig. 2, Tab. S6)."

R2C6. page 11: " … or could be a result of the sparsity of ELVOC data in the training set." More ELVOCS are available in Besel et al J. Aerosol Sci. 179, 106375 (2024): https://www.sciencedirect.com/science/article/pii/S0021850224000429?via%3Dihub

We appreciate the notice, however, the model we are referring to in this statement is one trained exclusively on experimental data. As far as we are aware, Besel et al. only offers quantum-mechanical calculation-derived vapor pressure data. We have also briefly tested the cross-application of models for experimental and quantum mechanical-derived vapor pressures and found substantial systematic errors that would make the mixing of data sets very difficult.

R2C7. Section 3.3. GC^2NN-GeckoQ: Do the authors have any insight into why GeckoQ is so much harder to learn? GeckoQ is by far the largest dataset of the ones studied in this work, so should be easier to learn or the learning curve should reach similar errors at higher training data volumes.

We have discussed this question, but unfortunately, we don't have a conclusive response. We have also discussed this with other groups using GeckoQ as a machine learning data set, and despite the data set size, to our knowledge, no one has so far been able to achieve similarly low errors than for smaller,

experimental data sets. However, we agree that this is surprising, because it means that the outcomes from COSMO-RS are somehow unpredictable by machine learning means.

R2C8. The "Conclusions" section is more of a Discussion section and should be named as such, because there is no other section named discussion, but there should be.

We would like to point out that the section prior to "Conclusions" is called "Results and Discussion" and includes the discussion of the points summarized in this last section. We agree that just "Conclusions" may be misleading here and have therefore renamed this section "Summary and Conclusions", as a concise summary of the most important aspects from the discussion section is followed by an elaboration of related research papers, future applications, remaining research questions and problems, and general ideas in the context of QSAR modelling.

R2C9. "… making spatial relations between molecular substructures directly interpretable." Did the authors try to interpret the models and extract chemical insight?

We would like to thank the reviewer for pointing out this potential misunderstanding. With this sentence, we tried to point out that the graph representations enhance a suitable machine learning model's capability of interpreting the molecular structure to derive chemical properties. For clarification, we have adapted the statement as follows:

> *l. 324 – "Graph representations are a natural and unambiguous representation of molecular structures, encoding additional information related to individual atoms (graph nodes) or bonds (graph edges), and making spatial relations between molecular substructures directly interpretable by ML models suitable for graph processing."*

In addition to our discussion around the molecular corridors of atmospheric oxidation that can be visualized with the output of our models (Fig. 5), we have now conducted a feature attribution analysis with regards to functional groups and elements, which can be regarded as chemical interpretation specific to the functionality of graph attention layers, see also our response to point R1C5.

References

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., and Pascanu, R.: Relational inductive biases, deep learning, and graph networks, https://doi.org/10.48550/ARXIV.1806.01261, version Number: 3, 2018.

Bilde, M., Barsanti, K., Booth, M., Cappa, C. D., Donahue, N. M., Emanuelsson, E. U., McFiggans, G., Krieger, U. K., Marcolli, C., Topping, D., Ziemann, P., Barley, M., Clegg, S., Dennis-Smither, B., Hallquist, M., Hallquist, A. M., Khlystov, A., Kulmala, M., Mogensen, D., Percival, C. J., Pope, F., Reid, J. P., Ribeiro da Silva, M. A. V., Rosenoern, T., Salo, K., Soonsin, V. P., Yli-Juuti, T., Prisle, N. L., Pagels, J., Rarey, J., Zardini, A. A., and Riipinen, I.: Saturation Vapor Pressures and Transition Enthalpies of Low-Volatility Organic Molecules of Atmospheric Relevance: From Dicarboxylic Acids to Complex Mixtures, Chem. Rev., 115, 4115–4156, https://doi.org/10.1021/cr5005502, 2015.

Burns, J., Zalte, A., and Green, W.: Descriptor-based Foundation Models for Molecular Property Prediction, https://doi.org/10.48550/ARXIV.2506.15792, version Number: 1, 2025.

Lansford, J. L., Jensen, K. F., and Barnes, B. C.: Physics-informed Transfer Learning for Out-of-sample Vapor Pressure Predictions, Propellants Explo Pyrotec, 48, e202200 265, https://doi.org/10.1002/prep.202200265, 2023.

Sanchez-Lengeling, B., Wei, J., Lee, B., Reif, E., Wang, P., Qian, W., McCloskey, K., Colwell, L., and Wiltschko, A.: Evaluating Attribution for Graph Neural Networks, in: Advances in Neural Information Processing Systems, edited by Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., vol. 33, pp. 5898–5910, Curran Associates, Inc., 2020.

Tang, B., Kramer, S. T., Fang, M., Qiu, Y., Wu, Z., and Xu, D.: A self-attention based message passing neural network for predicting molecular lipophilicity and aqueous solubility, Journal of cheminformatics, 12, 1–9, 2020.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y.: Graph Attention Networks, https://doi.org/10.48550/ARXIV.1710.10903, version Number: 3, 2017.

Withnall, M., Lindelöf, E., Engkvist, O., and Chen, H.: Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction, Journal of cheminformatics, 12, 1, 2020.

Zhang, Y.-J., Khorshidi, A., Kastlunger, G., and Peterson, A. A.: The potential for machine learning in hybrid QM/MM calculations, The Journal of Chemical Physics, 148, 241 740, https://doi.org/10.1063/1.5029879, 2018.